②

REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

AD-A250 603

to average 1 hour per response, including the time for reviewing instructions, searching existing data sources gathering and ion. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions ctorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to t and Budget, Washington, DC 20503.

REPORT DATE	3. REPORT TYPE AND DATES COVERED
1989	Unknown

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Probabilistic Inference and Non-Monotonic Inference	DAAB10-86-C-0567

6. AUTHOR(S)

Henry E. Kyburg, Jr.

DTIC
ELECTE
MAY 2 7 1992
S C D

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Rochester Department of Philosophy Rochester, NY 14627	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES).	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
U.S. Army CECOM Signals Warfare Directorate Vint Hill Farms Station Warrenton, VA 22186-5100	92-TRF-0011

1. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Statement A; Approved for public release; distribution unlimited.	

13. ABSTRACT *(Maximum 200 words)*

Since the appearance of the influential article by McCarthy and Hays, few people have tried to use probabilities as a basis for non-monotonic inference. One reason, perhaps the main one, is that probabilistic inference easily yields inconsistent bodies of knowledge, as is revealed by the lottery paradox. Here we establish three things: First that standard systems of non-monotonic reasoning (default logic, non-monotonic logic, and circumscription) fall prey to the same lottery-like difficulties as does probabilistic inference. Second, that probabilistic inference provides equally plausible treatment of the standard examples of non-monotonic reasoning. Third, that the inconsistency threatened by the lottery paradox is a petty hodgoblin, and need not in any way interfere with the use of beliefs in planning and design.

1. SUBJECT TERMS	15. NUMBER OF PAGES
Artificial Intelligence, Data Fusion, Commonsense Reasoning, Probability	12
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

②

REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

AD-A250 603

‖‖‖‖‖‖‖‖‖‖‖

to average 1 hour per response, including the time for reviewing instructions, searching existing data sources gathering and ion. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions ctorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to x and Budget, Washington, DC 20503.

REPORT DATE	3. REPORT TYPE AND DATES COVERED
1989	Unknown

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Probabilistic Inference and Non-Monotonic Inference	DAAB10-86-C-0567

6. AUTHOR(S)

Henry E. Kyburg, Jr.

DTIC
ELECTE
MAY 2 7 1992
S C D

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Rochester Department of Philosophy Rochester, NY 14627	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
U.S. Army CECOM Signals Warfare Directorate Vint Hill Farms Station Warrenton, VA 22186-5100	92-TRF-0011

1. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Statement A; Approved for public release; distribution unlimited.	

13. ABSTRACT (Maximum 200 words)

Since the appearance of the influential article by McCarthy and Hays, few people have tried to use probabilities as a basis for non-monotonic inference. One reason, perhaps the main one, is that probabilistic inference easily yields inconsistent bodies of knowledge, as is revealed by the lottery paradox. Here we establish three things: First that standard systems of non-monotonic reasoning (default logic, non-monotonic logic, and circumscription) fall prey to the same lottery-like difficulties as does probabilistic inference. Second, that probabilistic inference provides equally plausible treatment of the standard examples of non-monotonic reasoning. Third, that the inconsistency threatened by the lottery paradox is a petty hodgoblin, and need not in any way interfere with the use of beliefs in planning and design.

1. SUBJECT TERMS		15. NUMBER OF PAGES
Artificial Intelligence, Data Fusion, Commonsense Reasoning, Probability		12
		16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298, (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
299-01

Probabilistic Inference
and
Non-Monotonic Inference*

by
Henry E. Kyburg, Jr.
University of Rochester
Rochester, New York 14627

Commonsense Reasoning

Since the appearance of the influential article by McCarthy
and Hayes, few people have tried to use probabilities as a basis
for non-monotonic inference. One reason, perhaps the main
one, is that probabilistic inference easily yields inconsistent bodies
of knowledge, as is revealed by the lottery paradox. Here we
establish three things: First, that standard systems of
non-monotonic reasoning (default logic, non-monotonic logic, and
circumscription) fall prey to the same lottery-like difficulties as
does probabilistic inference. Second, that probabilistic inference
provides equally plausible treatments of the standard examples of
non-monotonic reasoning. Third, that the inconsistency
threatened by the lottery paradox is a petty hobgoblin, and need
not in any way interfere with the use of beliefs in planning and
design.

Probabilistic Inference
and
Non-Monotonic Inference

1. <u>Introduction.</u>

(1)   I have enough evidence to render the sentence S probable.

(2)   So, relative to what I know, it is rational of me to believe S.

(3)   Now that I have more evidence, S may no longer be probable.

(4)   So now, relative to what I know, it is not rational of me to believe S.

This seems like a perfectly ordinary, common sense, pair of situations. Generally and vaguely, I take them to embody what I shall call *probabilistic inference.* This form of inference is clearly non-monotonic. Very few people in AI seem to have looked at it carefully.

There are exceptions: Jane Nutter (1987) thinks that sometimes probability has something to do with non-monotonic reasoning. Judea Pearl has recently (1987) been exploring the possibility.

Of course there are any number of people whom one might call probability enthusiasts. Cheeseman (1985), Henrion (1987) and others think it useful to look at a distribution of probabilities over a whole algebra of statements, to update that distribution in the light of new evidence, and to use the latest updated distribution of probability over the algebra as a basis for planning and decision making. A slightly weaker form of this approach is captured by Nilsson (1986), where one assumes certain probabilities for certain statements, and infers the probabilities, or constraints on the probabilities, of other statements.

None of this corresponds to what I call probabilistic inference. All of the inference that is taking place is strictly

deductive. Deductive inference, particularly that concerned with the distribution of classical probabilities or chances, is of great importance. But as non-monotonists (?) have pointed out, this is not to say that there is not also an important role for what logicians have called "ampliative logic" or "inductive logic" or "scientific inference" -- that is, logic in which the conclusion goes beyond the premises, asserts more than do the premises. This is what David Israel (1980) has called "real rules of Inference." It is characteristic of any such logic or inference procedure that it can go wrong: that statements accepted at one point may be rejected at a later point.

2. McCarthy and Hayes.

As a matter of historical conjecture, I would suggest that it is the enormously influential article by John McCarthy and Pat Hayes (1969) that slowed the exploration of probabilistic inference. They offer powerful arguments against the use of probability as an approach to non-monotonicity. Since much of this argument applies equally well to most formalizations of non-monotonic inference, it is worth rehearsing their objections.

They commence with a first order predicate calculus, and add 3 operators: Consistent($\emptyset$), Normally($\emptyset$), Probably($\emptyset$). We consider a set σ of sentences, and add new sentences to it according to the rules:

1. Any consequence of σ may be added
2. If $\emptyset$ is consistent with σ, consistent($\emptyset$) may be added.
3. Normally($\emptyset$), Consistent($\emptyset$)$\vdash$ Probably($\emptyset$)
4. $\emptyset \vdash$ Probably($\emptyset$) is a possible deduction
5. If $\emptyset_1, \ldots, \emptyset_n \vdash \emptyset$ is a possible deduction, then

Probably($\emptyset_1$), $\ldots$, Probably($\emptyset_n$) $\vdash$ Probably($\emptyset$) is a possible deduction.

Two objections are offered against the use of probabilities:

"It is not clear how to attach probabilities to statements containing quantifiers in a way that corresponds to the amount of conviction people have." (p. 490)

"The information necessary to assign numerical probabilities is not ordinarily available. ... Therefore ... epistemologically inadequate." (p. 490)

2

A propos of these objections we only remark that (1964), and Krauss (1966), and others have provided schemes for assigning probabilities to quantified statements; and it is not obvious what information is necessary to assign numerical probabilities, so it's not obvious that we don't have it.

More telling is the example that McCarthy and Hayes offer: P looks up Q's number;

So he knows it;

So he should believe that if he dials it, "he will come into conversation with Q."

In order to obtain this conclusion in terms of probability, we need to add to the rules:

6.   Probably$(\emptyset) \vdash \emptyset$

We also need a rule for deleting statements from σ. But the details of these rules are not to the point. It is rather the intuitions that lie behind them (if not the axioms themselves) that have guided work in non-monotonic inference.

Rules (1) and (5) constitute the main stumbling blocks in the way of developing a system to embody these intuitions, even without (6). In particular, rule (1) requires that σ be strictly consistent; we shall argue later that this is neither necessary nor desirable.

From (1), (3), (4) and (5) and natural assumptions we get a lottery paradox (Kyburg 1961). (With (6) it becomes a contradiction.)

Assume: $\emptyset_i$: ticket i loses; $\emptyset$: all tickets lose; σ contains "$\sim\emptyset$"; Normally$(\emptyset_i)$; Consistent$(\emptyset_i)$;

By (3), Probably$(\emptyset_i)$. By (1) $\{\emptyset_i\} \vdash \emptyset$. Therefore by (5), Probably$(\emptyset)$. But by (4), Probably$(\sim\emptyset)$. If this isn't contradictory enough, add (6). Note that nothing self-referential is involved.

3. Non-monotonic Inference.

In this section we will show that three well-known examples of efforts to codify non-monotonic inference stumble equally over the lottery paradox. Before doing so, it will be well to consider the question of what we want of this set σ of

3

"believed" statements to do for us. Statements on that list should serve as _evidence_ for other statement. They should be useful as premises in _planning_ (the agent P is constructing a reasonable plan for talking to Q). They should define the limits (at a given point in the collection of evidence) of _reasonable_ or _serious possibility_. It is the planning and designing function of σ that is of clearest importance in AI. Don Perlis (1980) has come up with the term "use-beliefs" to describe this set of sentences. Perlis has also argued (Perlis 1980) that standard systems of non-monotonic reasoning have difficulty in handling iterations of application. His zookeeper example -- one of the birds is sick and can't fly -- is just the lottery paradox made realistic. But Perlis still seems to want σ to be consistent.

Let us look at the system of Reiter (1980), not because it is the latest thing in default systems, but because it embodies the intuitions driving default logic particularly clearly.

The intuitive idea is to adopt a set of defaults D inducing an extension E of some underlying incomplete set of wffs W. The extension E need not be unique, but (1), (2) and (3) should be satisfied:

(1) $W \subseteq E$,

(2)   E is deductively closed, and

(3)   Suppose $(A:MB_1, \ldots .MB_n/C)$ is a default, and each of B_i is consistent with E. Then C belongs to E.

The lottery with n tickets: W = {sentences describing a fair lottery}; a perfectly natural default is $:M{\sim}\emptyset_i/\emptyset_i$ – if you don't know that ticket i wins, i.e., if it is possible that it loses, it's reasonable to believe it loses.

This gives you n extensions E_i, each of which specifies one winner and n−1 losers. That seems implausible; it is clearly not useful for planning to have a number of extensions. Note that it is deductive closure that prevents our having a single extension containing each $\emptyset_i$.

How does circumscription handle the lottery tickets? Roughly speaking: If x is a ticket, and x is not abnormal, then x loses the lottery. (Of course if you don't know how many tickets there are in the lottery, it might be the case that circumscription would lead you to conclude that there was only

4

one, and that therefore you would win!) It is hard to know
exactly what one can conclude, since (McCarthy 1980) we
"clearly" have to include domain dependent heuristics for deciding
what circumscriptions to make and when to take them back. But
waiving difficulties of quantification, it certainly seems right that
in the case of just one ticket, we should be able to conclude that
it does not win the lottery.

The question then is, "How often can we iterate this
argument?" It would seem that the answer is roughly $n/2$: If we
have considered less than half the tickets, the next ticket is still
(statistically) likely to lose, relative to what we have taken
ourselves to know.

The non-monotonic logic of McDermott and Doyle (1980)
construes the set of non-monotonic theorems as (roughly) the
smallest fixed point under non-monotonic derivability. So it
would seem that in lottery cases there would be no useful set of
non-monotonic conclusions. And yet the lottery should not be
dismissed as frivolous: consider the lottery as standing for any
situation in which a certain outcome is taken to be "incredible",
and consider an arbitrarily long sequence of such situations. We
shall consider a specific case shortly.

4. The Cannonical Examples.

Before considering the seriousness of the inability of
standard non-monotonic inference systems to deal with the
lottery, let us show that probabilities do work reasonably well on
the standard examples of non-monotonic argument.

We assume that probabilities depend on our knowledge of
frequencies or chances, and further than they depend on what
we know about the object at issue.

Let E be evidence, K be the set of acceptable beliefs.
(1) Tweety.

> We know "almost all birds fly" in E.
> We know "all penguins are birds" in E.
> We know "no penguins fly" in E.

Suppose we add to E "Tweety is a bird." The probability
that Tweety flies is high. High enough for us to accept "Tweety
flies" in our body of knowledge (or use-beliefs) K.

Suppose we now add: "Tweety is a penguin" to E. The
probability that Tweety flies, relative to E is now 0. We must

5

<u>delete</u> "Tweety flies." And add "Tweety doesn't fly."

What we are doing here is using the <u>more specific</u> <u>reference class</u> as the basis for our probability. This principle is also given by McCarthy (1986), Etherington (1987)(as 'inferential distance'), Poole (1985) ("strictly more specific), and others.

(2) Nixon Diamond:

Add "Nixon is a Quaker" to the standard E. It is then probable relative to E that Nixon is a pacifist; we add "Nixon is a Pacifist" to K -- "we plan on it."

Now add "Nixon is a Republican" to the standard E. Similar: we add "Nixon is not a pacifist" to K.

Now add both. Knowledge about republicans and knowledge about quakers now conflict as a basis for asserting (or not) that Nixon is a pacifist. And we just don't know about the intersection. So we can't conclude anything. This conforms to the usual non-monotonic treatment. McCarthy (1986, p. 90) seems to give us a little more: we can conclude that Nixon is either an abnormal quaker or an abnormal republican. But the cash value of that is just that we don't know whether or not he is a pacifist.

(3) Cohabitation: (Reiter 1980)

Spouse(x,y)&hometown$(y)=z$:M Hometown$(x) = z$ /
hometown$(x) = z$

Employer(x,y)&Location$(y)=z$:M hometown$(x) = z$ /
hometown$(x) = z$

Consider John, whose spouse lives in Toronto and whose employer is located in Vancouver. We can derive an extension in which his hometown is Toronto, and one in which his hometown is in Vancouver.

Probability lets us conclude the disjunction: John's hometown is either Toronto or Vancouver. This may be false, of course, but that is the nature of probabilistic inference. The example suggests to Reiter: (1) ordering defaults (on what basis?) and (2) getting more information (a copout if we have to use what we evidence we have; otherwise, available to probability.

Reiter (1980) also suggests that default logic is useful for dealing with the frame problem. We use a default that says everything stays same unless it is deducible that it has changed. Probability does even better here: When we add a new statement

to E, it will not change most of the probabilities from which K is derived. So without a default: everything in K stays there unless it's _probability_ has changed.

This brings up the most serious charge against probable inference. Since what is probable depends on our whole body of evidence E, any addition to E requires the recomputation of all the probabilities on which K is based. But non-monotonic approaches require the same thing: "networks must be reconditioned after each update" (Touretzky (1986); Etherington (1987).) How expensive that reconditioning is depends on how difficult it is to compute probabilities.

5. Consistency.

How can K (the σ of McCarthy and Hayes) serve as a standard of serious possibility, as a basis for planning or designing, if it is inconsistent? Clearly, if it is deductively closed, it cannot. But it membership in K is determined by high probability it _won't_ be deductively closed. On the natural assumption that statements known to have the same truth value will have the same probability, it will contain (in principle) the logical consequences of statements that it contains. (Since if P entails Q, P ≡ P&Q will be logically true and thus part of the evidence E; and the probability of Q must be at least as great as that of P&Q.)

Inconsistency is a good reason for avoiding deductive closure. But what is a good reason for having an inconsistent K? The lottery is frivolous. Is there a practical and natural counterpart? The answer is yes, and it comes from the eminently respectable scientific domain of measurement. Suppose my job is measuring items produced by a certain machine. They are either OK or they aren't. It is clear that however sophisticated my measurements, it cannot be demanded that they be error free. Furthermore, there is no way of bounding measurement error. What _can_ be demanded is that of each measurement, the probability that it leads to a false assessment of OK be less than (say) .001. (Pick your own number if you don't like that one. But not 0.0.) For purposes of further manufacture, for purposes of design, for purposes of planning, it is obvious that I must believe, of each inspected piece, that it is

OK.

It is also clear that I ought also to believe -- accept as a serious possibility, accept for planning or design -- that of a large number (10^6?) items, at least one will not be OK. Thus my beliefs -- my use-beliefs -- are flatly inconsistent. The deductive closure of K contains all the statements in the language. Yet I have no difficulty using this (unclosed!) set of statements K for planning, or as a standard of serious possibility. (Even though the conjunction of statements in K is impossible, it serves as a standard of serious possibility in the sense that if a statement contradicts a member of K, it is not to be regarded as a serious possibility. That one of the pieces that has passed my inspection is not OK is not a serious possibility, though of course it is a possibility. And it is not a serious possibilty that all the inspected pieces are OK.)

The probabilistic version of non-monotonic inference faces one irrefutable complaint. Real probability enthusiasts will say that the story I have just told, while reflecting the way in which people do talk, should be regarded simply as a rough approximation to the real truth which can only be represented by a probability distribution over all the states of the world (or all the sentences of the language). In particular, in planning to use one of the manufactured items for a certain purpose, what I must "really" be doing is evaluating the probability that it is OK, given that it has passed my inspection, multiplying that by the utility of using a piece that really is OK, and adding to that the negative expectation, similarly computed, of using a piece that is not OK. If the probability of a false acceptance were .25, and the number of pieces involved were 10, clearly the probabilistic analysis would be preferable. Why not when the probabilities are smaller and the numbers bigger?

To this one can reply by waving one's hands at the computations involved: acceptance into the body of beliefs K is not only realistic (as a representation of what humans do) but is the only computationally feasible way to go about these things. But that answer is not clearly correct. Here it seems to me that we will only be able to get good answers about which is the best way to represent practical knowledge or use-beliefs by constructing systems that embody each approach.

6. Conclusions.

We have not provided an algorithm for performing probabilistic inference, though some efforts have been made along these lines (Loui 1986). We have not argued that pure probabilism is unworkable in principle. We have argued that many of the usual arguments against probabilistic inference are equally applicable to other forms of non-monotonic reasoning. We have argued that, in fact, once the single hurdle of "inconsistency" is overcome, probabilistic inference offers advantages (in some contexts) or at least no disadvantages (in most contexts) compared to non-monotonic reasoning. We have argued that, in fact, the ability to live comfortably with certain sorts of inconsistency is an important feature of probabilistic inference, and that it allows us to take as a basis for planning exactly those uncertainties that are collectively impossible to credit.

All this is not to say that the various forms of non-monotonic reasoning that have been explored are not useful for special purposes. It does suggest both that specific instances of non-monotonic argument can be justified by reference to probabilities, and that any sort of inference that was incompatible with probabilistic inference would have some strikes against it.

references:

Cheeseman, Peter (1985): "In Defense of Probability," *IJCAI 85*, Morgan Kaufmann, Los Altos, 1002-1009.

Etherington, D. W. (1987): "Formalizing Non-Monotonic Reasoning Systems," *Artificial Intelligence* 31, 41-86.

Gaifman, Haim (1964): "Concerning Measures on First Order Calculi," *Israeli Journal of Mathematics* 2, 1-18.

Henrion, Max, and Cooley, Daniel (1987): "An Experimental Comparison of Knowledge Engineering for Expert Systems and for Decision Analysis," *AAAI 87*, Morgan Kaufman, Los Altos, 471-476.

Israel, David (1980): "What's Wrong with Non-Monotonic Logic?" *IJCAI 1980*, Stanford, 99-101.

Kyburg, Henry E., Jr. (1961): *Probability and the Logic of Rational Belief*, Wesleyan University Press, Middletown.

Loui, Ronald P. (1986): "Computing Reference Classes," *Proceedings of the 1986 Workshop on Uncertainty in Artificial Intelligence*, 183-188.

McCarthy, John (1980): "Circumscription -- A Form of Non-Monotonic Reasoning," *Artificial Intelligence* 13, 27-39.

McCarthy, John (1986): "Applications of Circumscription to Formalizing Common-Sense Knowledge," *Artificial Intelligence* 28, 89-116.

McCarthy, John, and Hayes, Pat (1969): "Some Philosphical Problems from the Standpoint of Artificial Inteligence," Machine Intelligence 4, 1969, 463-502, reprinted in Weber and Nilsson (eds) *Readings in Artificial Intelligence*, Tioga Publishing, Palo Alto, 1981.

McDermott, D., and Doyle, J. (1980): "Non-Monotonic Logic I," *Artificial Intelligence* 13, 41-72.

Nilsson, Nils (1986): "Probabilistic Logic," Artificial Intelligence 28, 71-87.

Nutter, Jane Terry (1987): "Uncertainty and Probability," *IJCAI 87*, Morgan Kaufmann, Los Altos, 373-379.

Pearl, Judea, and Geffner, Hugo (1987): "

Perlis, Donald (1986): "On the Consistency of Commonsense Reasoning, "

Poole, David L. (1985): "On the Comparison of Theories: Preferring the Most Specific Explanation," *IJCAI 85*, 144-147.

Reiter, R. (1980): "A Logic for Default Reasoning," *Artificial Intelligence* 13, 81-132.

Scott, Dana, and Krauss, Peter (1966): "Assigning Probabilities to Logical Formulas," in Hintikka and Suppes (eds) *Aspects of Inductive Logic*, North Holland, Amsterdam,

219-264.

Touretzky, D. S. (1986): *The Mathematics of Inheritance Systems*, Morgan Kaufman, Lost Altos, 1986.